

NLP for Crowdfunding Success Prediction: A Systematic Mapping Review Exposes an Underexplored Equity Domain

Chaya Liebeskind¹ Hana Tzur² Shmuel Liebeskind³

¹Department of Computer Science, Jerusalem College of Technology

²Department of Industrial Engineering and Management, Jerusalem College of Technology

³Department of Information Science and Applied Artificial Intelligence, Bar-Ilan University

{liebchaya, hanatzur6, israellieb}@gmail.com

Abstract

Equity crowdfunding under U.S. Regulation Crowdfunding (Reg CF) raised over \$1.4 billion across 5,969 offerings between May 2016 and December 2022 (Chang, 2024), generating a large public corpus of Form C business descriptions on SEC EDGAR. Yet regulated equity crowdfunding remains barely studied by the financial NLP community. We conduct a systematic mapping review using automated database search, independent dual-LLM screening, targeted human adjudication, and blind quality-control auditing. We screened 726 deduplicated papers and included 49 studies applying NLP directly to crowdfunding campaign text. Only one study examines U.S. Reg CF campaign text directly, using readability, with one pre-Reg-CF Title II near-neighbor using sentiment tone and description length. The available equity-crowdfunding evidence is therefore too limited and heterogeneous to support general conclusions about the strength of textual signals in regulated equity settings. We taxonomize the included studies into five NLP method families, synthesize representative cross-study evidence in the original reported units, and identify the near absence of regulated equity research across the methodological spectrum. We also discuss regulatory constraints, structured financial information, and text length as plausible but untested explanations for this gap. The annotated evidence base, screening records, and reproducibility materials are publicly available at <https://github.com/jct-nlp/finNLP-crowdfunding>.

1 Introduction

Equity crowdfunding under U.S. Regulation Crowdfunding (Reg CF) has matured into a major startup financing channel, raising over \$1.4 billion cumulatively across 5,969 offerings between May 2016 and December 2022 (Chang, 2024). Every issuer must file a standardized Form C with the

SEC, creating a massive, public corpus of corporate descriptions on EDGAR. Read by retail and accredited investors to evaluate high-stakes offerings, these descriptions sit at the intersection of regulatory disclosure and persuasive pitch-making.

Despite the scale of this data, financial NLP has almost entirely overlooked regulated equity crowdfunding. Research has focused nearly exclusively on reward platforms (Kickstarter, Indiegogo) and donation sites (GoFundMe), where campaign narratives may differ in length and structure and are not subject to the disclosure requirements that apply to regulated securities offerings.

Prior work shows that campaign-text and presentation signals are associated with crowdfunding outcomes (Mollick, 2014). Text-based representations have also been shown to contribute to predictive models (Pekar et al., 2025; Ye et al., 2024). These findings are scattered across platforms and evaluation protocols. Existing reviews synthesize crowdfunding success factors broadly and within specific crowdfunding regimes (Deng et al., 2022; Mochkabadi and Volkmann, 2020; Hohen et al., 2025; Alegre and Moleskis, 2021). However, to our knowledge, no prior review has systematically mapped NLP representations applied directly to campaign-level text across crowdfunding models, compared their evaluation designs, and isolated the evidence available for U.S. Regulation CF.

We address this gap using a structured, auditable systematic mapping review pipeline. We screened a 726-paper corpus and identified 49 studies applying NLP to campaign text. Our analysis reveals a stark domain gap: textual features are measurably predictive on reward- and donation-based platforms, yet evidence from regulated equity crowdfunding remains scarce. Only one of the 49 included studies applies NLP to campaign text under Reg CF proper (Akter et al., 2026), alongside a single pre-Reg-CF Title II near-neighbor (Bjornsen et al., 2024). The available evidence from these two studies is

too limited and heterogeneous to support a general claim about textual signals in equity settings.

Our contributions are: (1) An auditable AI-assisted evidence-synthesis workflow combining automated retrieval, independent dual-LLM screening, targeted human adjudication, and quality-control auditing; (2) An organizing taxonomy of NLP methods used in crowdfunding research; (3) A cross-study evidence synthesis; (4) Empirical characterization of the equity research gap, together with hypothesized (but empirically untested) explanations for why textual signals might differ in regulated equity offerings; and (5) A curated annotated evidence base and reproducibility materials released with this paper.

2 Methodology

Beyond supporting the present review, the screening workflow provides a practical framework for large-scale evidence synthesis. Rather than replacing human reviewers, the pipeline combines independent LLM judgments, selective human adjudication, and post-screening quality control to reduce manual screening effort while retaining explicit decision traces. We executed the systematic mapping review using automated database search, dual independent LLM screening, quality-control audit sampling, and human adjudication (Figure 2). The review characterizes how NLP methods have been applied to crowdfunding research, with particular attention to textual sources, crowdfunding models, prediction tasks, methodological choices, and underexplored settings.

2.1 Search Phase & Query Logging

Our review targeted publications from January 2010 through December 2025 using CrossRef, Semantic Scholar, OpenAlex, arXiv, and Google Scholar. We used the following core query:

(crowdfunding OR crowd-funding) AND
(NLP OR text mining OR natural language OR sentiment OR topic model OR BERT OR language model OR readability OR linguistic)

The displayed query represents the shared search logic. Complete database-specific queries, search dates, targeted equity-search queries, and search logs are provided in the accompanying public repository. Records were deduplicated first by DOI and then by title similarity using a threshold of

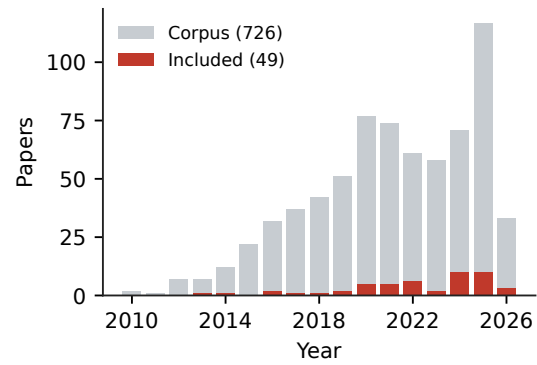


Figure 1: Year distribution of the 726-paper corpus and the 49 included studies (2010–2025, updated through June 2026, with three 2026 studies).

0.85; no formal language or publication-type filters were imposed.

The initial search was conducted in September 2025, followed by targeted equity-crowdfunding updates in 2026 using platform- and regulation-specific queries (e.g., Reg CF, the JOBS Act, Wefunder, StartEngine, and EU/UK equity platforms). Google Scholar was not re-queried during these updates. These 2026 updates were intentionally equity-focused rather than symmetric updates across all crowdfunding regimes, because their purpose was to increase search sensitivity for the equity-focused research question. Accordingly, the 2026 records should not be interpreted as a complete cross-regime census of crowdfunding research for that year. After cross-source deduplication, the corpus contained 726 papers: 625 from the initial search and 101 from the 2026 update rounds, which also captured additional late-2025 records and three studies published in 2026. Because the corpus was assembled iteratively, we use the deduplicated total of 726 as the retrieval anchor (Figure 2). The included studies are concentrated in recent years, with three from 2026 (Figure 1).

2.2 Screening Phase & Dual-LLM Review

To scale screening with scientific rigor, we deployed a dual-LLM pipeline. Two advanced, independent models—GPT-5.4-mini (via Azure) and Gemini 3.5 Flash—evaluated each of the 726 abstracts independently and blindly. For each paper, the screening record contained three criterion-level responses (C1–C3, defined below), an overall label in {INCLUDE, EXCLUDE, UNCERTAIN}, a 1–3 sentence justification, and a confidence score (0.0 to 1.0). The complete screening prompt, output

schema, and available model/deployment configuration details are provided in the public repository.

Inclusion rules. A paper was classified as `INCLUDE` if and only if: (R1) it relates to crowd-funding; (R2) it applies NLP, text mining, or language modeling; (R3) it reports at least one quantitative outcome metric; and (R4) NLP is applied directly to campaign-level text (e.g., descriptions, updates, transcripts). Otherwise, it was labeled `EXCLUDE`. The screening tool operationalizes these four rules as three fields— $C1 = R1$; $C2 = R2 \wedge R4$, i.e., NLP applied specifically to campaign-level text, evaluated as a single joint check rather than two separate yes/no fields; $C3 = R3$ —so all references below to the combined R2/R4 criterion refer to this joint field C2. To operationalize the combined R2/R4 criterion, we verified that the reported NLP methods were applied directly to campaign-level text rather than to external textual sources (e.g., news articles or social media).

Agreement & Adjudication. Decisions were resolved under two cases:

- **Case A (Full Agreement):** Both models agreed on `INCLUDE` or `EXCLUDE`. The decision was automatically accepted.
- **Case B (Disagreement / Uncertainty):** Any label mismatch or `UNCERTAIN` call was escalated to human adjudication.

A human expert reviewed all escalated cases, acting as the final ground truth. Across all 726 papers, the two LLMs returned a concordant, definitive decision on 580 (79.9%, Case A, auto-accepted); the remaining 146 (20.1%, Case B—a label mismatch or an `UNCERTAIN` vote by either model) were escalated to human adjudication. Chance-corrected agreement on the three-way label is moderate (Cohen’s $\kappa = 0.590$), but the models agree substantially on the substantive criterion of whether NLP is applied to campaign text ($\kappa = 0.787$, which falls within the “substantial” range proposed by Landis and Koch (1977)). Of the 146 escalated cases the human included 41 and excluded 105 (Figure 3). The lower agreement on the overall three-way screening label primarily reflects uncertainty handling, whereas agreement on the substantive campaign-text inclusion criterion (C2) is substantially higher. Operationally, concordant definitive decisions allowed 79.9% of records to bypass

initial human adjudication. These agreement statistics measure inter-model consistency rather than screening correctness or recall; screening validity is considered separately through the human audit described below.

2.3 Human-in-the-Loop Audit Sampling

As an independent quality check, we drew a blind sample of 87 Case A papers (15% of the 580 auto-accepted) and submitted them for blind human review without revealing the LLM consensus labels. To ensure consistent application of the predefined screening protocol, the reviewed records were subsequently re-examined against the inclusion criteria. Only one record required correction (1.15%; exact Clopper–Pearson 95% CI: 0.03%–6.2%).

The corrected record was the peer-to-peer lending study by Mezei et al. (2018), which the LLM consensus had initially accepted for inclusion. The study was excluded because its “linguistic data transformation” referred to fuzzy-set representations of structured variables rather than NLP or analysis of campaign-level text. The same exclusion decision was independently reproduced during the subsequent full-text extraction stage.

Thus, 86 of the 87 audited records (98.85%) were confirmed without requiring any change to the screening outcome. Because the audit sample was drawn from Case A records without stratification by consensus `INCLUDE` versus `EXCLUDE` decisions, it was not designed to estimate screening recall or the false-exclusion rate. Separately from this screening-stage audit, all 49 included studies also passed through a two-pass extract-and-verify pipeline over the full PDF text (Section 3), providing an additional full-text validation of the extracted study characteristics. The full-text stage used Claude Sonnet 5 in two sequential passes: an initial structured extraction followed by verification against the source PDF; ambiguous or conflicting central fields were resolved through human verification.

Together, these components provide an auditable hybrid screening and verification workflow that may be useful for future evidence-synthesis studies beyond crowdfunding.

2.4 Metadata Annotation Layer

Every included study was annotated with metadata describing the crowdfunding context. Studies were classified as Reward-based, Donation-based,

Reviewer pair	n	Raw agreement	Cohen's κ
<i>Overall three-way screening label (INCLUDE/EXCLUDE/UNCERTAIN)</i>			
M1 $\leftrightarrow$ M2	726	88.3%	0.590
<i>Combined R2/R4 campaign-text criterion</i>			
M1 $\leftrightarrow$ M2	726	97.5%	0.787

Table 1: Inter-rater agreement on the 726-paper screening corpus. M1 = GPT-5.4-mini via Azure. M2 = Gemini 3.5 Flash. The two LLMs returned a concordant, definitive decision on 580/726 papers (79.9%, Case A, auto-accepted); the remaining 146 (20.1%, Case B—label mismatch or an UNCERTAIN vote) were escalated to human adjudication. Raw agreement on the three-way label includes records for which both models returned UNCERTAIN, whereas Case A comprises only concordant, definitive INCLUDE/EXCLUDE decisions. Chance-corrected agreement on the three-way screening label is moderate (Cohen’s $\kappa = 0.590$), but the two LLMs agree substantially on the substantive criterion of whether NLP is applied to campaign text ($\kappa = 0.787$, “substantial” on the Landis and Koch (1977) scale). A blind human QC audit of 87 Case A papers (15% of the 580 agreed) resulted in one correction after re-examination against the predefined inclusion criteria (1.15%).

Equity crowdfunding, or Peer-to-peer (P2P) lending. For the cross-study analysis, equity-crowdfunding studies were further distinguished by regulatory context as non-U.S. equity, pre-Reg-CF U.S. Title II, or U.S. Reg CF. Two additional categories were used when necessary: Multiple crowdfunding models for studies spanning more than one crowdfunding model, and Other for funding mechanisms outside the taxonomy adopted in this review.

Platform, crowdfunding-model, and regulatory-context labels were manually verified against the full texts. Studies spanning multiple crowdfunding models were classified as *Multiple crowdfunding models*, whereas funding mechanisms outside the taxonomy adopted in this review were classified as *Other*. Every inclusion decision is fully traceable to either dual-LLM consensus or explicit human adjudication.

3 Taxonomy of NLP Methods

We organize the 49 included studies into five families of text representation (Figure 4). These families are intended as an organizing synthesis of the representations observed in the corpus rather

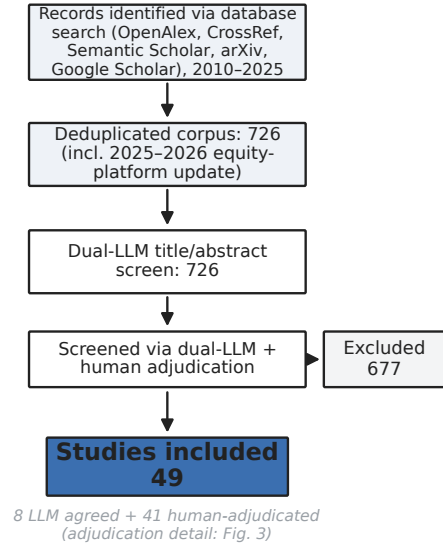


Figure 2: Study identification and selection flow. The 726-paper deduplicated corpus was screened by a dual-LLM pipeline with human adjudication, yielding 49 included studies. Adjudication detail in Figure 3.

than as a novel domain-specific NLP taxonomy. Figure 5 further shows how these method families are distributed across crowdfunding regimes. The descriptions below draw on deep reads of selected illustrative papers. The initial NLP-family labels were generated automatically using keyword-based rules run over each study’s verified full-text extraction (the linguistic-information and NLP-representation fields captured by our two-pass extract-and-verify pipeline over the full PDF text, not just the abstract). All labels were subsequently verified manually against the full texts, and disagreements or missing labels were resolved through human adjudication, yielding a complete family assignment for all 49 studies, compared with the ~ 22 an abstract-only pass would have left unclassified. The families follow a roughly chronological progression from classical lexical representations to modern generative and LLM-based approaches.

Figure 4 shows the overall prevalence of each NLP family. Sentiment analysis (53.1%) and semantic/topic representations (51.0%) are the most frequently used approaches, followed by lexical methods (46.9%) and readability measures (28.6%). LLM-derived methods remain rare (6.1%), reflecting the very recent adoption of LLMs within the crowdfunding literature.

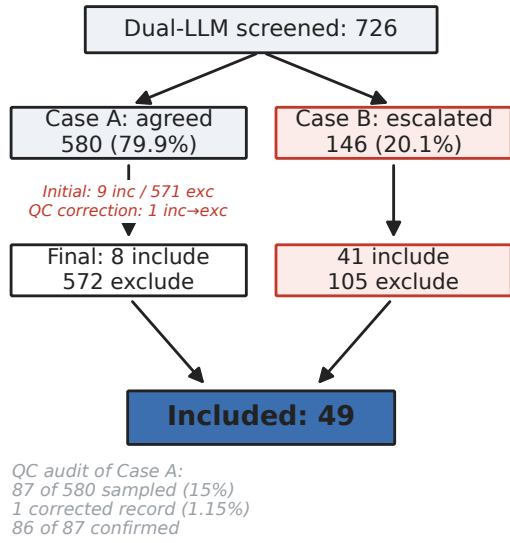


Figure 3: Adjudication and audit. Both LLMs screened all 726 abstracts; 580 (79.9%) received concordant decisions and were automatically accepted, whereas 146 (20.1%) were escalated to human adjudication. A blind quality-control audit of 87 auto-accepted (Case A) papers resulted in one corrected screening decision after re-examination against the predefined inclusion criteria, changing the initial Case A consensus from 9 INCLUDE / 571 EXCLUDE to a final 8 INCLUDE / 572 EXCLUDE. The corrected record was a peer-to-peer lending study that did not apply NLP or textual analysis to campaign-level text. The final corpus therefore comprises 8 Case A studies and 41 human-adjudicated Case B studies (49 total).

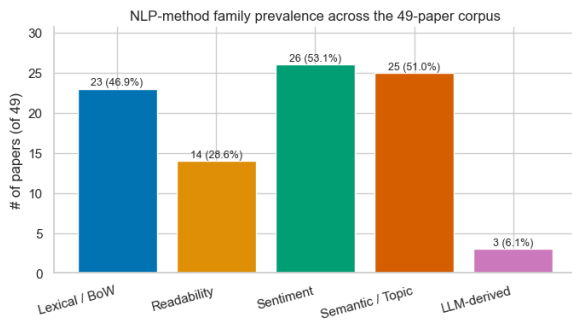


Figure 4: Prevalence of NLP method families across the 49 included studies. Percentages are calculated over the included corpus. A study may contribute to multiple families.

The heatmap highlights a strong methodological imbalance across crowdfunding regimes. Nearly every NLP family has been applied predominantly to reward-based crowdfunding, with donation-based studies representing a secondary concentration. In contrast, regulated equity crowdfunding

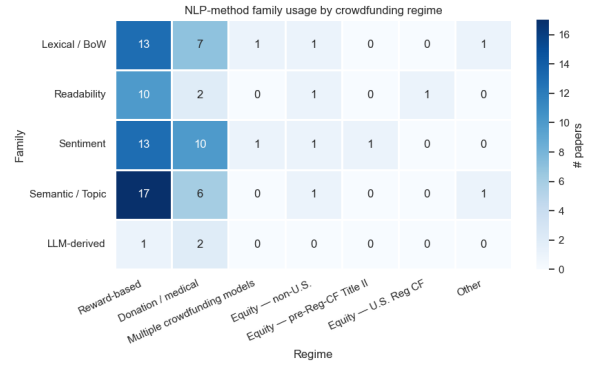


Figure 5: Distribution of NLP method families across crowdfunding regimes. Each cell reports the number of included studies using a given NLP family within a crowdfunding regime. Counts are not mutually exclusive because individual studies may employ multiple NLP method families.

contains only isolated examples across all method families, indicating that the research gap extends beyond individual techniques to the overall methodological landscape.

Family 1: Lexical and bag-of-words. Early crowdfunding prediction studies used small sets of handcrafted textual features, including sentiment, readability, and sentence count (Greenberg et al., 2013). Lexical and bag-of-words methods represent text using word-frequency lists (TF-IDF and keyword counts) or dictionaries such as LIWC (Pennebaker et al., 2001).

Most studies in this family analyze the CAMPAIGN_DESCRIPTION directly. For instance, von Selasinsky and Isaak (2020) used LIWC to analyze 1,049 Kickstarter project descriptions and video transcripts. Their text-only logistic-regression models achieved McFadden’s pseudo- R^2 values of 0.289 for project descriptions, 0.117 for YouTube transcripts, and 0.109 for Otter transcripts.

Family 2: Readability and linguistic style. Instead of vocabulary, these methods measure *how* a campaign communicates. This includes readability scores (Flesch, Gunning Fog) and style markers (first-person plural, certainty language, hedging, future orientation). For example, Akter et al. (2026) measure the readability of Form C business descriptions (CAMPAIGN_DESCRIPTION) across 285 U.S. Reg CF campaigns and find a positive coefficient associated with funding success ($\beta = 0.056, p < .05$), one of the few readability findings available for a regulated equity setting.

Scaled by the reading-ease score’s sample SD of 8.94 points, this corresponds to roughly 65% higher odds of funding success per one-SD increase in readability—a scale-appropriate reading of the coefficient, not a magnitude the source paper states directly.

Family 3: Sentiment analysis. These papers use sentiment dictionaries (TextBlob, VADER) or machine learning to capture the emotional tone of descriptions, on the general premise that a more positively-toned pitch reads as more persuasive to backers. However, general-purpose sentiment tools may struggle with financial language, motivating domain-adapted models such as FinBERT (Araci, 2019), which was further pretrained on Reuters financial news and evaluated on financial sentiment datasets.

Family 4: Semantic and topic representations. These methods use tools like LDA (Blei et al., 2003) or BERTopic (Grootendorst, 2022) to capture overall themes rather than individual words. For example, Pekar et al. (2025) compared RAKE-extracted keyword features, contextual topic models (built on DistilRoBERTa), and sentence clustering (using Universal Sentence Encoder) across 267 Kickstarter 3D-printing campaigns. Under random forest, contextual topic models and keyword lists reduced RMSE relative to the median baseline by 10.12% and 8.85%, respectively, while sentence clusters reduced it by 4.75% (not significant). Under LASSO, sentence clusters reduced RMSE by 6.23% (significant), slightly more than contextual topic models (5.90%). Overall, Pekar et al. (2025) found no reliable evidence that one representation consistently outperformed the others, although sentence clusters were easier for human raters to interpret.

Family 5: LLM-derived features. The newest studies leverage instruction-tuned LLMs to extract semantic and structural themes directly from unstructured text. For example, Ye et al. (2024) used GPT-4 (gpt-4-1106-preview) to extract 11 semantic and pragmatic narrative features from 11,274 GoFundMe campaigns. Their best all-feature LightGBM model, combining non-textual, GPT-derived, and lexicon-based textual features, achieved 81.0% test accuracy, compared with 59.3% for the paper’s uniform majority-class baseline, representing a 36.6% relative improvement. In a separate ablation analysis, a LightGBM model us-

ing only non-textual features also achieved 59.3% accuracy; adding the 11 GPT-derived features increased accuracy to 73.9%. The identical 59.3% values therefore refer to two distinct comparisons. In a counterfactual simulation, GPT-4 was used to revise campaign descriptions by adding expressions of gratitude, explaining the urgency of the funding need, and acknowledging the availability of matching grants. These revisions increased the model-predicted funding probability from 33.2% to 45.1%, an increase of 11.9 percentage points.

Figure 5 further shows that this methodological evolution has occurred primarily within reward-based crowdfunding, with only isolated examples in regulated equity settings. This suggests that the equity research gap is not confined to a particular NLP technique but extends across the full methodological spectrum represented in the literature.

Platform distribution and the equity gap. The included corpus is dominated by unregulated reward-based platforms, followed by donation/medical crowdfunding; all remaining crowdfunding settings are represented by only a small number of studies (Figure 6). Critically, **only one of the 49 included studies applies NLP to campaign text under U.S. Regulation CF proper**: Akter et al. (2026) examine the readability of Form C business descriptions (CAMPAIGN_DESCRIPTION) across 285 StartEngine and WeFunder campaigns and report a positive coefficient associated with funding success ($\beta = 0.056, p < .05$); in robustness checks against continuous funding outcomes (log funds raised, funding speed), every other variable in their model remained significant, but readability specifically was not. The closest comparator is Bjornsen et al. (2024), who model equity-crowdfunding success using sentiment tone and campaign-description length in pre-2016 Title II offerings. Both textual variables reached significance only at the 10% level in a balanced-training specification, no non-text-only baseline was reported, and the study predates Reg CF; we therefore treat it as a near-neighbor rather than a Reg CF study.

The evidence therefore indicates that regulated equity crowdfunding remains only minimally studied rather than entirely unstudied. Only one Reg CF study and one adjacent pre-Reg-CF Title II study analyze campaign text, and the available evidence is too limited and heterogeneous to support general conclusions about textual signals in regu-

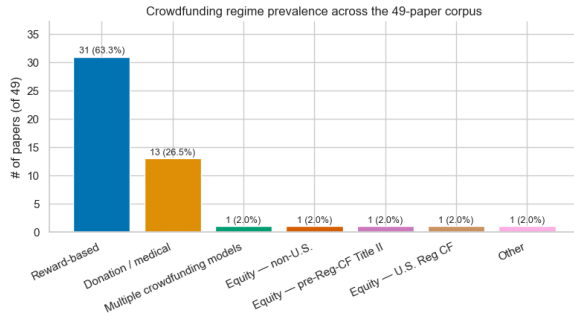


Figure 6: Distribution of the included studies across crowdfunding regimes, highlighting the concentration of NLP research in reward-based crowdfunding and the scarcity of regulated equity studies.

lated equity crowdfunding. This gap motivates the cross-study synthesis presented next.

Before synthesizing the cross-study evidence, we note that two studies could not be assigned to a single one of the four primary crowdfunding-model categories. The study by [Zheng et al. \(2025\)](#) analyzed both reward- and equity-based platforms and was therefore classified as *Multiple crowdfunding models*. The study by [Mansouri and Momtaz \(2022\)](#) examined blockchain-based utility-token initial coin offerings and was classified as *Other*. These studies remain in the corpus but are not analyzed as evidence for any single model.

4 Cross-Study Evidence

An illustrative comparison of representative studies is provided in Appendix A (Table 2). The table is intended to illustrate the diversity of reported outcomes, evaluation protocols, and evaluation metrics across the major crowdfunding settings rather than to enable direct comparison of predictive performance. Table 2 also illustrates a recurring reporting limitation: in several representative studies, the incremental contribution of text cannot be isolated because a comparable non-textual baseline or ablation is not reported. Here, we focus on synthesizing the principal methodological patterns observed across the included studies.

Key observations. Two broad patterns emerge from the representative studies summarized in Appendix A (Table 2). First, several reward-crowdfunding studies report that textual features contribute to prediction, but the magnitude of this contribution cannot be compared directly because the studies use different outcomes, baselines, metrics, and evaluation designs. One possible expla-

nation is that textual characteristics differ across crowdfunding settings, although this was not examined directly in the review. Second, the available equity-crowdfunding studies provide limited and methodologically heterogeneous evidence regarding the incremental predictive contribution of campaign text: [Akter et al. \(2026\)](#) find a statistically significant readability effect on whether a campaign passes or fails under Reg CF proper, and [Bjornsen et al. \(2024\)](#) report tone and word-count coefficients that reach significance only at the 10% level in a balanced-training specification, with no non-text-only baseline reported. With only these two studies we cannot generalize about text effects in equity crowdfunding; the primary finding is a research gap rather than evidence that textual signals are inherently weak in equity crowdfunding.

5 Discussion

Beyond its substantive findings, this study demonstrates a practical workflow for AI-assisted evidence synthesis that combines independent LLM screening with targeted human verification. The combination of independent LLM reviewers, human adjudication, and targeted quality-control auditing enabled the screening of hundreds of papers while maintaining transparent decision traces and explicit quality assessment. Similar workflows may be applicable to other large-scale evidence-synthesis tasks in NLP and computational social science. The taxonomy also reveals that the methodological gap is not limited to a particular NLP technique. Instead, nearly every major NLP family has been explored almost exclusively in reward-based crowdfunding, while regulated equity crowdfunding remains largely absent across the methodological spectrum.

Possible explanations for the equity gap. The limited presence of equity crowdfunding in the NLP literature may reflect underlying differences between regulated equity and reward-based crowdfunding rather than simply research attention. Several factors may help explain why the contribution of campaign text remains uncertain in equity crowdfunding. Regulatory standardization, the availability of structured financial information, differences in document format, and variation in investor decision-making may affect the usefulness of textual signals. However, the small and heterogeneous evidence base does not allow these explanations to be tested in the present review; we offer three

plausible mechanisms as untested hypotheses:

1. *Rules and legal constraints.* Regulated equity offering documents may undergo legal review, and issuers face incentives to avoid overstated or potentially misleading promotional claims. We hypothesize that such constraints could reduce some of the lexical and rhetorical variation exploited by sentiment and readability models. This mechanism remains untested in the present review.
2. *Availability of structured financial information.* Equity investors can draw on structured signals such as equity retention and detailed risk disclosures, which have been associated with funding success (Ahlers et al., 2015). We therefore hypothesize that, when such signals are available, linguistic tone may provide less incremental predictive value. In Kickstarter and other reward-based settings, by contrast, backers may rely more heavily on the information conveyed through the campaign page, although this contrast has not been systematically established.
3. *Potential differences in text length.* We hypothesize that differences in document length could affect the relative utility of different text representations, although this relationship was not tested in the present review.

These three hypothesized factors offer one possible explanation for the equity research gap identified in this review, though our corpus cannot distinguish between them. These hypotheses should not be interpreted as evidence that textual effects are weak in equity crowdfunding. They also suggest that researchers looking for language signals in equity crowdfunding should look at other channels: Form C-U update filings (CAMPAIGN_UPDATES), transcripts of pitch videos (VIDEO_TRANSCRIPT), or social media buzz (SOCIAL_MEDIA) around the campaign.

Open challenges. The NLP crowdfunding research space still has three major hurdles to overcome. First, we need to know how well models generalize across platforms. No study has tested whether a model trained on Kickstarter works on Indiegogo, let alone transfers to an equity platform. Since each community has its own slang and style, we cannot just assume models will work everywhere. Second, we lack public datasets for Reg

CF research. Although Form C filings are public, reliable outcome labels may require supplementing incomplete Form C-U filings with platform or third-party data. Creating an open, shared dataset would help more researchers enter this space. Third, there are ethical and legal risks to using AI to rewrite pitches. While using GPT-4 to rewrite GoFundMe pages shifted the model’s predicted probability of funding success by 11.9 percentage points in one study (Ye et al., 2024), doing the same for regulated equity offerings could raise legal and disclosure-compliance concerns if the AI invents details or misrepresents the business.

6 Conclusion

We presented a systematic mapping review of NLP methods for crowdfunding success prediction. From a 726-paper deduplicated corpus, the dual-LLM screening workflow identified 49 studies applying NLP directly to campaign text, spanning five methodological families. The literature is concentrated in reward- and donation-based crowdfunding, and reported effects are not directly comparable because studies use different outcomes, baselines, metrics, and evaluation designs.

Methodologically, the review illustrates an auditable evidence-synthesis workflow combining independent LLM screening with selective human adjudication and blind quality control. The two LLMs produced concordant definitive decisions for 79.9% of the screened corpus, while the blind audit of auto-accepted records resulted in one corrected screening decision. These results demonstrate a substantial reduction in the number of records requiring initial human adjudication while preserving explicit auditability. They should not, however, be interpreted as an estimate of screening recall or as evidence that false exclusions were eliminated.

Substantively, we identify a near-total gap in regulated U.S. equity crowdfunding: only one included study directly examines Reg CF campaign text, with one pre-Reg-CF Title II near-neighbor. The available evidence is too limited and heterogeneous to support general conclusions about textual effects in equity settings. The primary finding is therefore a lack of evidence rather than evidence that textual signals are inherently weak in equity crowdfunding. Regulatory constraints, structured financial information, and text length remain plausible but untested explanations, while the public availability of Form C filings makes Reg CF a

promising setting for future financial NLP research.

Data Availability

The complete evidence table covering all 49 included studies, together with the search logs, study manifest, screening annotations, extraction records, coding schema, manually verified taxonomy labels, screening prompt, and reproducibility materials, is publicly available at <https://github.com/jct-nlp/finNLP-crowdfunding>. Appendix A provides an illustrative cross-study comparison of selected studies.

Limitations

Our review has several limitations. First, any systematic mapping review is constrained by its retrieval strategy. We may have missed grey literature, workshop papers, non-English studies, or work applying textual analysis without standard NLP terminology. In addition, the 2026 update was intentionally targeted toward equity crowdfunding and did not re-query Google Scholar; it should therefore not be interpreted as a complete, symmetric update across all crowdfunding regimes. Second, screening decisions were based on titles and abstracts. Although a blind audit of 87 Case A records resulted in one corrected decision, the audit was not stratified by consensus INCLUDE versus EXCLUDE decisions and does not provide an estimate of screening recall or false-exclusion rates. Relevant studies incorrectly excluded at this stage may therefore remain undetected. Third, the remaining 146 escalated cases were adjudicated by a single human expert. Fourth, the taxonomy depends on a predefined coding schema and human resolution of ambiguous cases; because studies may combine several representations, some family assignments retain interpretive subjectivity despite full-text verification. Fifth, the 49 included studies differ substantially in crowdfunding model, textual source, outcome definition, evaluation protocol, and reported metric, preventing formal meta-analysis and direct performance ranking. Finally, we focused on studies reporting quantitative outcomes and therefore excluded purely qualitative work.

References

Gerrit KC Ahlers, Douglas Cumming, Christina Günther, and Denis Schweizer. 2015. Signaling in equity

crowdfunding. *Entrepreneurship Theory and Practice*, 39(4):955–980.

Mohsina Akter, Tyge-F. Kummer, and Peter Green. 2026. How issue-relevant information and peripheral cues shape investment decisions in equity crowdfunding. *Electronic Commerce Research*, 26(3):2605–2641.

Inés Alegre and Melina Moleskis. 2021. Beyond financial motivations in crowdfunding: A systematic literature review of donations and rewards. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 32(2):276–287.

Dogu Araci. 2019. FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Matt Bjornsen, Sarah Borchers, Bree Dority, and Suzanne Hayes. 2024. Predicting equity crowdfunding success: An examination of united states offerings using sentiment analysis. *Journal of Accounting, Business and Management (JABM)*, 31(1):65–79.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Melody Chang. 2024. Women and minority-owned businesses in regulation crowdfunding. Technical report, U.S. Securities and Exchange Commission, Office of the Advocate for Small Business Capital Formation. Cumulative Regulation CF capital raised, May 2016–December 2022.

Lingfei Deng, Qiang Ye, Dapeng Xu, Wenjun Sun, and Guangxin Jiang. 2022. A literature review and integrated framework for the determinants of crowdfunding success. *Financial Innovation*, 8(1):41.

Michael D. Greenberg, Bryan Pardo, Karthic Hariharan, and Elizabeth M. Gerber. 2013. Crowdfunding support tools: predicting success & failure. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, pages 1815–1820.

Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Sascha Hohen, Christian Hüning, and Lars Schweizer. 2025. Reward-based crowdfunding—a systematic literature review. *Management Review Quarterly*. Published online 11 August 2025.

J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Sasan Mansouri and Paul P. Momtaz. 2022. Financing sustainable entrepreneurship: ESG measurement, valuation, and performance. *Journal of Business Venturing*, 37(6):106258.

- József Mezei, Ajay Byanjankar, and Markku Heikkilä. 2018. Credit risk evaluation in peer-to-peer lending with linguistic data transformation and supervised learning. In *Hawaii International Conference on System Sciences*, pages 1366–1375. Proceedings of the Annual Hawaii International Conference on System Sciences.
- Kazem Mochkabadi and Christine K. Volkmann. 2020. [Equity crowdfunding: a systematic review of the literature](#). *Small Business Economics*, 54(1):75–118.
- Ethan Mollick. 2014. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, 29(1):1–16.
- Viktor Pekar, Marina Candi, Ahmad Beltagui, Nikolaos Stylos, and Wei Liu. 2025. Explainable text-based features in predictive models of crowdfunding campaigns. *Annals of Operations Research*, 354(1):367–397.
- James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. *Linguistic Inquiry and Word Count: LIWC2001*. Erlbaum, Mahwah, NJ.
- Constantin von Selasinsky and Andrew Jay Isaak. 2020. It’s all in the (sub-)title? expanding signal evaluation in crowdfunding research. In *Proceedings of the 28th European Conference on Information Systems (ECIS 2020)*.
- Teng Ye, Jingnan Zheng, Junhui Jin, Jingyi Qiu, Wei Ai, and Qiaozhu Mei. 2024. Using artificial intelligence to unlock crowdfunding success for small businesses. *arXiv preprint arXiv:2407.09480*.
- Yilong Zheng, Yiru Wang, and Sarfraz A. Mian. 2025. [Strategic positioning of projects in crowdfunding platforms: do advanced technology terms referencing, signaling and articulation matter?](#) *International Journal of Entrepreneurial Behavior & Research*, 31(2-3):751–780.

A Illustrative Cross-Study Comparison

This appendix provides an illustrative comparison of representative studies discussed in the main text.

Study / ID	Platform (<i>n</i>)	Family	Outcome	Model	Eval. design	Metric	Result	Incr. text contribution
Pekar et al. (2025)	KS 3D-print (267)	Semantic Topic	/ Log-transformed pledged amount	Random Forest	10-fold metrics averaged across folds	CV; RMSE (log scale)	2.09 (−10.1% vs. median baseline)	n.r. (text-only vs. median-predictor baseline; no covariate model reported)
Ye et al. (2024)	GoFundMe (11,274)	LLM-derived	Pred. prob. of funding success	LightGBM	Counterfactual GPT-4 rewrite simulation	Change in pred. prob.	+11.9 pp (simulated)	n/a—a simulated intervention effect, not an observed outcome
von Selasinsky and Isaak (2020)	KS (1,049)	tech Lexical BoW	/ Binary funding success	Logistic regression	Nested model comparison	McFadden pseudo- R^2 (as reported)	0.289 (description-only model)	Δ McFadden pseudo- $R^2 \approx 0.025$ (Model 1 controls-only: 0.871 → Model 3 controls + description: 0.896)
Akter et al. (2026)	StartEngine / Wefunder (285)	Readability	Binary funding success (Reg CF)	Logistic regression	Single model + robustness checks	Unstd. logit β (Flesch RES, SD=8.94)	$\beta = 0.056$, $p < .05$	n.r. (no non-text-only baseline reported)
Bjornsen et al. (2024)	U.S. Title II pre-Reg CF (3,216)	Sentiment	Binary funding success	Logistic regr. (combined text+non-text model)	Balanced ($n=64$) vs. orig. imbalanced training	Logit coef.	$p < .10$ only (balanced); n.s. (imbalanced)	cannot be isolated (no ablation or text-only baseline)

Table 2: Illustrative comparison of representative studies spanning the major crowdfunding settings discussed in this review. The table is intended as a qualitative cross-study reference rather than a direct quantitative comparison because the studies differ substantially in their prediction tasks, outcomes, baselines, evaluation protocols, and reported metrics. Abbreviations: KS = Kickstarter; BoW = bag-of-words; RES = Reading Ease Score; pp = percentage points; n.r. = not reported; n/a = not applicable; n.s. = not significant. For [von Selasinsky and Isaak \(2020\)](#), the unusually high pseudo- R^2 values for the control-containing models are reported directly by the source; the control set includes the log of the amount pledged.